UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/488,028	01/20/2000	Eric Cohen-Solal	US 000015	3849

24737 7590 04/18/2005

PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P.O. BOX 3001
BRIARCLIFF MANOR, NY 10510

EXAMINER	
HAILU, TADESSE	
ART UNIT	**PAPER NUMBER**
2173	

DATE MAILED: 04/18/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/488,028	COHEN-SOLAL ET AL.	
	Examiner	Art Unit	
	Tadesse Hailu	2173	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06 December 2004*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,2,4 and 6-15* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) *12* is/are allowed.

6)☒ Claim(s) *1,2,4,6-10 and 13-15* is/are rejected.

7)☒ Claim(s) *11* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is in response to REPLY entered on December 6, 2004 for the patent application (09/488,028), filed on 1/20/2000.

2. The pending claims 1, 2, 4, and 6-15 are hereby examined.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1, 2, 4, 6, 7, and 13-15 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tom Brøndsted, et al "The IntelliMedia WorkBench A Generic Environment For Multimodal Systems," (1998) in view of Christopher R. Wren, et al. "Combining Audio and Video in Perceptive Spaces," December 13-14, 1999.

With regard to claim 1:

As per "a method of locating and displaying an image of a target," Brondsted describes a method of locating and displaying an image of a target (see fig. 1).

As per "sensing a triggering event generated by a human operator;" Brondsted describes sensing spoken word (key word or command) as well as user's gesture via a microphone and camera respectively (see section 3);

As per "receiving additional external information that characterizes at least one machine-sensible feature of a target, said receiving step occurring substantially simultaneously with said sensing step;" since Brondsted is a multimodal system, thus additional information about a target or location can be received through spoken word (extracted key word) input as well as through gesture input (section 3). These inputs are executed simultaneously (section 2.1).

Brondsted also discloses that the sensing step includes sensing a gesture, such as a pointing gesture (see Fig.1, sections 1 and 2) indicting a direction of said target. Furthermore, Brondsted discloses directing or aiming a camera toward a target (Fig. 1), but Brondsted does not, however, discloses directing or aiming the camera toward the target in response to said sensing and said receiving step.

Wren, on the other hand, describes a Perceptive Spaces applying to specific application, such as for example City of News (section 3.3). In this section, as in Brondsted's *workbench*, Wren also describes SMART DESK, wherein to navigate the City of News, virtual 3D, users sit in front of the SMART DESK an uses voice and hand gestures to explore or load new data (see section 3.3, Fig. 7). In regard to claimed subject matter, Wren further describes coupling of gesture and speech modalities to redirect/move camera to the desired target (see section 3.3, page 5). As described in this section the user of the system points to a link (target of interest) and says "there" to

load a new URL page, in response the virtual camera will automatically move to a new position in space that constitutes an ideal view point of the current page. Thus, Wren discloses aiming a camera in response to said sensing (e.g. hand gesture) and receiving (e.g. keyword or command word/speech) steps as specified in the claim.

Brondsted and Wren are analogous art because they are from the same field of endeavor that is multi modal system.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to modify the ceiling mounted camera (fixed view, Fig. 1) camera of Brondsted by substituting for a swivel or movable camera of Wren so that the it can be directed to a target in response to gesture and speech input of the user as described by Wren (section 3.3, 2nd column).

The suggestion/motivation for doing so would have been to provide optimal viewpoints and constrained navigation so that the user is never lost in the virtual world (section 3.3, 2nd column)

Therefore, it would have been obvious to combine Brondsted with Wren to obtain the invention as specified in claim 1.

With regard to claim 2:

As per "... said sensing step includes sensing a gesture of a human operator indicating a target." Brondsted in view of Wren discloses Gesture recognizer (fig. 2) for sensing a gesture of a human operator indicating a target (see Brondsted, fig. 1).

With regard to claim 4:

As per "... said receiving step includes receiving speech from said human operator." Brondsted in view of Wren discloses Microphone (fig. 2) for receiving speech from said human operator (see Brondsted, section 2.1).

With regard to claim 6:

As per "... processing said speech for use with at least one machine sensor, said at least one machine sensor and said speech assisting in locating said target." Brondsted in view of Wren disclose Speech recognizer, Speech synthesizer, and Microscope (see Brondsted, fig. 2, and section 2.1).

With regard to claim 7:

As per "... said sensing step includes sensing a gesture indicting a direction from said human operator to said target." Brondsted in view of Wren discloses a gesture indicating a direction form said human operator to said target (see Brondsted, fig. 1).

With regard to claim 13:

As per "A method of aiming a camera at a target," Brondsted illustrates aiming a camera and a laser pointer at a campus map location (target) (fig. 1).

As per "inputting an indication of a position of a target;" Brondsted illustrates and describes pointing toward a location of a target (fig. 1, see also section 3);

As per " inputting further information about a machine-sensible characteristic of said target;" Brondsted describes sensing spoken word (key word or command) as well as user's gesture via a microphone and camera respectively (see section 3);

Brondsted further discloses that said inputting an indication step includes inputting a gesture (Fig.1. sections 2-2.1) indicating a direction of said target.

Furthermore, Brondsted discloses aiming a camera toward a target (Fig. 1), but Brondsted does not, however, discloses directing or aiming the camera toward the target in response to said indication and said further information as required in claim 13.

Wren, on the other hand, describes a Perceptive Spaces applying to specific application, such as for example City of News (section 3.3). In this section, as in Brondsted's *workbench*, Wren also describes SMART DESK, wherein to navigate the City of News, virtual 3D, users sit in front of the SMART DESK and uses voice and hand gestures to explore or load new data (see section 3.3, Fig. 7). In regard to claimed subject matter, Wren further describes coupling of gesture and speech modalities to redirect/move camera to the desired target (see section 3.3, page 5). As described in this section the user of the system points to a link (target of interest) and says "there" to load a new URL page, in response the virtual camera will automatically move to a new position in space that constitutes an ideal view point of the current page. Thus, Wren discloses aiming a camera in response to said sensing (e.g. hand gesture) and receiving (e.g. keyword or command word/speech) steps as specified in the claim.

Brondsted and Wren are analogous art because they are from the same field of endeavor that is multi modal system.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to modify the ceiling mounted camera (fixed view, Fig. 1) camera of Brondsted by substituting for a steering or movable camera of Wren so that the it can be directed to a target in response to gesture (indication) and speech input (other input or further information) of the user as described by Wren (section 3.3, 2nd column).

The suggestion/motivation for doing so would have been to provide optimal viewpoints and constrained navigation so that the user is never lost in the virtual world (section 3.3, 2nd column).

Therefore, it would have been obvious to combine Brondsted with Wren to obtain the invention as specified in claim 13.

With regard to claim 14:

As per "A method of acquiring a target," Brondsted illustrates a method of acquiring a target using a camera and a laser pointer within at a campus map environment for example, to locate office location/address within the campus (target) (Fig. 1, sections 2-2.1).

As per "inputting spatial information to indicate a position of a target" Brondsted illustrates (Fig.1, pointing) and describes pointing toward a location of a target (see also section 3).

As per " inputting further information about said target" Brondsted describes inputting spoken word (key word or command) as well as user's gesture via microphone and camera respectively (see section 3).

As per "spatial information includes sensing a gesture indicating a direction of said target" Brondsted as illustrated in Fig. 1and as described in section 2, discloses spatial information (pointing toward the target) includes sensing a gesture (the system through its sensors (e.g. camera) senses the gesture) indicating a direction of said target

But Brondsted does not discloses "orienting an instrument with respect to said target in response to said spatial information and said further information and said further information to reduce an ambiguity in said position"

Wren, on the other hand, describes a Perceptive Spaces applying to specific application, such as for example City of News (section 3.3). In this section, as in Brondsted's *workbench*, Wren also describes SMART DESK, wherein to navigate the City of News, virtual 3D, users sit in front of the SMART DESK and uses voice and hand gestures to explore or load new data (see section 3.3, Fig. 7). In regard to claimed subject matter, Wren further describes coupling of gesture and speech modalities to redirect/move camera to the desired target (see section 3.3, page 5). As described in this section the user of the system points to a link (target of interest) and says "there" to load a new URL page, in response the virtual camera will automatically move to a new position in space that constitutes an ideal view point of the current page. Wren further describes the coupling of gesture and speech modalities are used to avoid false recognitions or ambiguity (Wren, page 5, last paragraph). Thus, Wren discloses orienting camera (an instrument) with respect to said target in response to said user pointing (spatial information) and said speech (further information) steps as specified in the claim.

Brondsted and Wren are analogous art because they are from the same field of endeavor that is multi modal system.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to modify the ceiling mounted camera (fixed view, Fig. 1) camera of

Brondsted by substituting for a steering or movable camera of Wren so that the it can be directed to a target in response to gesture (pointing, spatial information) and speech input (other or further information) of the user as described by Wren (section 3.3, 2nd column).

The suggestion/motivation for doing so would have been to provide optimal viewpoints and constrained navigation so that the user is never lost in the virtual world (section 3.3, 2nd column).

Therefore, it would have been obvious to combine Brondsted with Wren to obtain the invention as specified in claim 14.

With regard to claim 15:

As per "...said step of orienting includes orienting a camera." Brondsted in view of Wren, as illustrated in fig. 1 of Brondsted, shows oriented camera view toward a workbench.

4. Claims 8-10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tom Brøndsted, et al "The IntelliMedia WorkBench A Generic Environment For Multimodal Systems," (1998) in view of Christopher R. Wren, et al. "Combining Audio and Video in Perceptive Spaces," December 13-14, 1999 further in view of Indrajit Poddar, et al "Toward Natural Gesture/Speech HCI: A Case Study of Weather Narration," 1998.

With regard to claim 8:

As per "...said processing step includes processing said voice information through a look-up table corresponding said speech to search criteria for use with said at

least one sensor." Brondsted in view of Wren describes different module for storing data, but Brondsted in view of Wren fails to describe, "processing said voice information through a look-up table corresponding to said speech to search criteria for use with said at least one sensor." Similar to Brondsted, Poddar discloses a multimodal system, including speech (via Microphone) and gesture (hand) input (section 3). Poddar, on the other hand, further discloses processing voice information through a look-up table (table1- table 4).

Brondsted, Wren and Poddar are analogous art because they are from the same field of endeavor that is multi-modal system.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to replace Brondsted's voice information memory storage with Poddar's look-up table because it would be easier to structure/formulate the voice information and access the voice information in a table format (see pages 3 and 5).

Therefore, it would have been obvious to combine Brondsted and Wren with Poddar to obtain the invention as specified in claims 8 through 10.

With regard to claim 9:

As per "... said look-up table is modifiable." Brondsted in view of Wren and Poddar further describe replacing key words of the table, modifiable look-up table (Poddar, section 3).

With regard to claim 10:

As per "...said look-up table modifiable by receiving information through the on-line global compute network." Since Brondsted in view of Wren and Poddar can be

implemented in a distributed environment (see Brondsted sections 2.1- 2.2), the look-up table (voice data memory module) could be modified by information received from other remote devices.

Allowable Subject Matter

5. Claim 12 is allowed.

The following is an examiner's statement of reasons for allowance: the prior art of records teaches all the steps recited in claim 12 except for "aiming a camera in response to said sensing, storing and said receiving steps."

Brondsted in view of Wren describes a simultaneous speech and gesture input implemented on Workbench (see section 2.1). Brondsted in view of Wren further describes and illustrates (fig. 1) a camera directed toward the target, wherein the camera continuously captures the pointing hand over the workbench while the user/operator describes the location (section 2.1). Furthermore, while Brondsted in view of Wren discloses for "aiming a camera in response to said sensing and said receiving steps, but Brondsted in view of Wren fails to disclose all the required limitations as recited above in claim 12.

Thus, prior art neither renders obvious nor anticipates the combination of claimed elements in light of the specification.

6. Claim 11 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The following is a statement of reasons for the indication of allowable subject matter: Although Brondsted and Poddar describes a modifiable look-up table (Poddar, section 3) that includes replaces word or phrase input with another input and a corresponding search criteria (Poddar, section 3), " said added voice input and said corresponding search criteria established by comparing previous association of said added voice input with at least one machine sensible characteristic of at least one correctly identified target associated with said voice input, said machine sensible characteristic being a basis for determining said corresponding search criteria." not clearly described.

Response to Arguments

7. Applicant's arguments filed December 6, 2004 have been fully considered but they are not persuasive.

The applicant argues that the Office action fails to show that Brondsted in view of Wren teaches or suggests the claim 1 recitation of *"receiving additional external information that characterizes at least one machine-sensible feature of a target*". In contrast to the Applicant's argument Brondsted in view of Wren teaches "*receiving additional external information that characterizes at least one machine-sensible feature of a target"* as recited in claim 1.

The Examiner maintains the same position in describing the above argument. That is, the claimed "additional external information" corresponds for example in a multi modal campus information system to a user's spoken enquiry about a specific office

location, location of a person, etc as claimed in claim 1. For example, when a user of the system is interested to find out once Office location, For example, Hanne's, the user simply asks "*show me Hanne's office*" through the disclosed microphone, the Workbench which comprises speech recognizer outputs or responds by saying (e.g., "T*his is Hanne's office.*" (Brondsted, section, 2.1).

Applicant further argues "there is no showing that Brondsted teaches or suggests that the words "Hanne" and/or "office" characterizes a feature of a target of the campus map of Fig. 1 that is detected by a sensor, for example."

In contrast to the applicant's argument Brondsted's multi modal campus information system characterizes the user's input made to the workbench table (campus information (Fig. 1) using the Software architecture or modules (Fig. 2) of the workbench. For example, the system allows the user to ask questions about the location of persons and offices, labs, etc, then the system analyzes the question or the spoken word (via one or more modules, Fig. 2) and outputs the intended output whether spoken (via speaker, "T*his is Hanne's office.*") or gestures (e.g., pointing coordinates). The system therefore receives additional external information that characterizes at least one machine-sensible features of a target.

Accordingly, the system receives inquires (e.g., "*show me Hanne's office")* from the user. The inquires, which are characteristic or attribute feature of a target are analyzed and/or compared (via one or more modules, Fig. 2) with the pre-stored campus information, and the system retrieves and outputs the answer whether spoken (via speaker, "T*his is Hanne's office.*") or gestures (e.g., pointing coordinates) to the

inquires. Again, Brondsted teaches "machine-sensible features of a target," such as Rooms are describes by an identifier for the room (room number) and the type of the room (office, corridor, etc). For Offices there is also a description of tenants of the room by a number of attributes (first and second name, affiliation, etc.) (Brondsted, section 3, see Domain Model). At least all these are "machine-sensible features of a target.

Accordingly, the Office action show that Brondsted in view of Wren teaches inputting (or receiving) additional external information that characterizes (e.g. Hanne's office) at least one machine-sensible feature (e.g. speech recognizer module is sensible to recognize key phrases, "Hanne", "office" and output the result, that is "This is Hanne's office").

The applicant also argues neither Brondsted nor Wren is cited for teaching or suggesting at least the claim 14 recitation of *"orienting an instrument with respect to said target to acquire said target in response to said spatial information and said further information to reduce an ambiguity in said position*".

In contrast to the Applicant's argument, Brondsted in view of Wren teaches coupling of gesture and speech modalities to redirect/move or "orienting" camera to the desired target (see Wren, section 3.3, page 5). As described in this section the user of the system points to a link (target of interest) and says "*there*" to load a new URL page, in response, the virtual camera will automatically moves or "orients" to a new position ("spatial information") in space that constitutes an ideal view point of the current page target (see Wren, section 3.3, page 5).

Accordingly, Wren discloses orienting camera (an instrument) with respect to said target in response to said user pointing (spatial information) and said speech (further information) steps as specified in the claim.

The applicant strongly argues there is no mention or showing of "to reduce an ambiguity in said position" of the target.

In contrast to the Applicant's argument, Brondsted in view of Wren teaches integrating or coupling of gesture and speech modalities are used to avoid false recognitions or ambiguity (Wren, page 5, last paragraph).

Furthermore, the Applicant argues that Poddar is not cited for curing any of the deficiencies of Brondsted and Wren described above with respect to claim 1. In contrast to the Applicant argument, the only claimed feature not shown in Brondsted and Wren is the "look up table". However, Poddar teaches the claimed (claims 8-10) "look-up table". Similar to Brondsted, Poddar discloses a multi-modal system, including speech (via Microphone) and gesture (hand) input (section 3). Poddar, on the other hand, further discloses processing voice information through a look-up table (table1- table 4).

Furthermore in regard to dependent claims 2, 4, 6, 7, and 15, the rest of the limitations found in these dependent claims are also described in the reference of record (see the Office action). They too are unpatentable.

Having fully addressed the applicant's arguments, the rejection still stands.

Conclusion

8. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

9. Any inquiry concerning this communication or earlier communications from the Examiner should be directed to Tadesse Hailu, whose telephone number is (571) 272-4051. The Examiner can normally be reached on M-F from 10:00 - 630 ET. If attempts to reach the Examiner by telephone are unsuccessful, the Examiner's supervisor, John Cabeca, can be reached at (571) 272-4048 Art Unit 2173.

10. An inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group receptionist whose telephone number is (703) 305-3900.

Tadesse Hailu

April 7, 2005